Arabinogalactan-Proteins: Key Regulators at the Cell Surface?

Miriam Ellis, Jack Egelund, Carolyn J. Schultz, and Antony Bacic*

Plant Cell Biology Research Centre, School of Botany (M.E., A.B.), and Australian Centre for Plant Functional Genomics (A.B.), University of Melbourne, Melbourne, Victoria 3010, Australia; Department of Plant Biology and Biotechnology, Faculty of Life Sciences, University of Copenhagen, DK–1871 Frederiksberg C, Denmark (J.E.); and School of Agriculture, Food, and Wine, University of Adelaide, Glen Osmond, South Australia 5064, Australia (C.J.S.)

Arabinogalactan-proteins (AGPs) are undoubtedly one of the most complex families of macromolecules found in plants, perhaps matched only by the phenolics (lignins/cutins/suberins) and pectins. Their complexity arises from the incredible diversity of the glycans decorating the protein backbone, the array of peripheral sugars decorating the large arabinogalactan (AG) chains, the microheterogeneity of protein backbone glycosylation, and the diversity of protein backbones containing AG glycomodules. It has been postulated that up to 40% of the Arabidopsis (Arabidopsis thaliana) proteins predicted to be glycosylphosphatidylinositol anchored have the potential to be substituted with AG chains; whether these are all AGPs remains to be established. This complexity is reminiscent of the mammalian extracellular glycoproteins/proteoglycans that are known to be critically important as both structural and functional determinants. These parallels have provided the impetus for plant scientists to establish roles for AGPs in plant growth and development as biological regulatory molecules, with some tantalizing possibilities emerging. They have also been widely used by society for industrial and food applications due to their general emulsifying, adhesive, and water-holding properties. Some AGs/AGPs, such as larch AG and gum arabic, have more recently been investigated as potential immunomodulators of the human immune system.

AGPs belong to a large family of wall glycoproteins/proteoglycans, the Pro/Hyp-rich glycoproteins, originally classified into three separate classes: Pro-rich proteins (PRPs), extensins, and AGPs (Showalter, 1993; Nothnagel, 1997). From various experimental and genomic analyses, it is clear that this family comprises a continuum of molecules from the nonglycosylated/minimally glycosylated PRPs, the moderately glycosylated extensins, to the highly glycosylated AGPs (for review, see Clarke et al., 1979; Fincher et al., 1983; Kieliszewski and Lamport, 1994; Nothnagel, 1997; Cassab, 1998; Serpe and Nothnagel, 1999; Gaspar et al., 2001; Showalter, 2001; Johnson et al., 2003b; Seifert and Roberts, 2007). Despite this continuum and the identification of chimeric and hybrid molecules, there are a large number of molecules that reflect the original classification of extensins, PRPs, and AGPs. This complexity is reminiscent of animal extracellular matrix glycoproteins/proteoglycans and reflects their varied roles as both key structural and regulatory molecules (Filmus et al., 2008; Schaefer and Schaefer, 2010).

AGPs are found on the plasma membrane, in the wall, in the apoplastic space, and in secretions (e.g. stigma surface and wound exudates). They have been found in detergent-resistant membranes in Arabidopsis (Arabidopsis thaliana), suggesting their presence in lipid rafts (Borner et al., 2005). Key distinguishing features (with notable exceptions) of AGPs appear in (1) their carbohydrate, primarily O-linked to the Hyp residues of the protein backbone, that constitutes 90% to 98% (w/w), usually as branched type II arabinino-3,6-galactans (AGs; 5–25 kD), although short oligoarabinosides are found (Fig. 1); (2) their protein that typically constitutes 1% to 10% (w/w) and is rich in Hyp/Pro, Ala, Ser, and Thr, with the dipeptide motifs Ala-Hyp, Ser-Hyp, Thr-Hyp, Val-Pro, Gly-Pro and Thr-Pro as distinguishing features, although Ser-(Hyp)_{2}-(extensin) motifs can also be present; (3) the presence of a glycosylphosphatidylinositol (GPI) membrane anchor, predicted on most, but not all, AGP protein backbones based on the presence of a hydrophobic, C-terminal domain in the proteins encoded by gene sequences (Fig. 1A); and (4) the ability of most, but not all, AGPs to bind a class of synthetic chemical dyes, the Yariv reagents, in particular the β-glucosyl Yariv reagent (Yariv et al., 1967), that has...
Figure 1. A, The wattle blossom model of the structure of AGPs with a GPI membrane anchor attached. In this model, there are approximately 25 Hyp residues. Most Hyp residues are noncontiguous and are predicted to bear an AG chain. Each AG chain may contain 15 or more repeats of a β-(1-3)-linked Gal oligosaccharide. There may be a few contiguous Hyp residues bearing short arabino-oligosaccharides. The molecule as a whole is spheroidal. The structure of the GPI anchor shows an ethanolamine-phosphate (P) between the anchor and the C terminus of the protein backbone, which is common to all GPI anchors. The core oligosaccharide of the GPI shown is based on PcAGP1 from pear (*Pyrus communis*; Oxley and Bacic, 1999), which comprises 2- and 6-linked Manp residues, a 4-linked GlcNH$_2$ residue, and a monosubstituted inositol with a partial Galp residue substitution to C(6) of the 6-linked Manp residue. The lipid moiety is a ceramide composed primarily of a phytosphingosine base and tetracosanoic acid. This model is modified from Fincher et al. (1983). B, The twisted hairy rope model of the structure of the GAGP. A hypothetical block size of 7 kD contains 10 amino acid residues (1 kD), 30 sugar residues (4.4
proven extremely useful in their detection, quantification, and precipitation from solution, a useful early step in their purification. Although the precise mechanism of Yariv binding is not known, it requires the presence of both carbohydrate and protein moieties and is highly variable in strength (Pettolino et al., 2006).

The high degree of heterogeneity displayed by AGPs is a property of the complexity of both the protein backbone gene families, gleaned largely from the genome sequencing programs, and their carbohydrate sequence and composition, revealed by a combination of chemical studies and the use of antibodies (Knox, 1997). These antibody studies and more recently molecular biology approaches using synthetic Hyp-rich motifs (Tan et al., 2004; Estevez et al., 2006) have also revealed this heterogeneity within and between tissue/cell types, sustaining the interest of researchers regarding their potential as biological regulatory molecules. Mounting evidence indicates that AGPs have a role in controlling plant growth and development. Although much of the evidence is circumstantial (for review, see Seifert and Roberts, 2007), GPI-anchored proteins in mammalian and yeast cells are involved in cell-cell signaling by interacting with other proteins, and the GPI anchor determines their location in membrane microdomains or membrane rafts enriched in sphingolipids and sterols (Orlean and Menon, 2007; Kinoshita et al., 2008). Thus, GPI anchors on AGPs provide a plausible mechanism by which AGPs may be involved in signaling pathways (Schultz et al., 1998; Johnson et al., 2003b). Additionally, it is clear that carbohydrate epitopes on mammalian, Drosophila melanogaster, and protozoan parasite proteoglycans/glycoproteins are key determinants of biological specificity/function (Haltiwanger and Lowe, 2004; Filmus et al., 2008; Schaefer and Schaefer, 2010). However, the key question remains as to whether the function of AGPs resides in the protein backbones, the exquisite microheterogeneity of the glycan epitopes, or both. In this review, we summarize the key structural features of AGPs and overview their putative functions. More exhaustive reviews on each of these topics are cited above.

STRUCTURE

Carbohydrate

The carbohydrate is usually in the form of polysaccharide chains, type II AGs that are O-glycosidically linked to Hyp residues on the protein backbone. These type II AGs have (1-3)-β-D-linked Galp residues that form a backbone substituted at C(O)6 by galactosyl side chains, usually terminating in Araf, Rhap, and Galp residues, giving rise to a neutral glycan chain (Fig. 1). Some AGs, such as those found in gum arabic, are also rich in GlcpA (with or without 4-O-methyl ether) residues and are primarily terminal, giving the molecules an overall negative charge. Native AG chains can range in size from 5 to 25 kD (i.e. approximately 30–120 sugar residues; Fincher et al., 1983; Gane et al., 1995). Chemical analyses of some AG chains suggests that the (1-3)-β-galactan backbone contains repeat blocks of approximately seven Galp residues interrupted by a periodate-sensitive linkage, possibly either (1-6)-Galp or (1-5)-Araf (Churms et al., 1981, 1983; Bacic et al., 1987). Short oligoarabinosides linked O-glycosidically to Hyp residues (Fig. 1) and single Gal residues O-glycosidically linked to Ser/Thr residues are also found on some AGPs (Goodrum et al., 2000).

Characterization of the carbohydrate structures of AGPs is complicated by the difficulties in isolating single AGP molecules and the microheterogeneity of the constituent glycan chains. The application of molecular biology techniques to isolate both heterologously expressed AGP protein backbones and synthetic peptides as GFP fusion proteins by the Kieliszewski/Showalter and Matsuoka laboratories (Shpak et al., 1999; Zhao et al., 2002; Shimizu et al., 2005) was an ingenious innovation and allowed direct testing of the Hyp contiguity hypothesis (Kieliszewski and Lamport, 1994) that predicts glycosylation patterns. Synthetic AGPs with single dipeptide repeats [e.g. (Ala-Hyp)n] reduced the microheterogeneity to a level permitting NMR-based sequencing of AG chains on the synthetic peptides (Fig. 1C), thereby providing major advances in our knowledge. The expression of a major tomato (Solanum lycopersicum) AGP (LeAGP1) as a fusion protein with GFP in tobacco (Nicotiana tabacum) BY-2 suspension culture cells facilitated its purification, and the Hyp glycoside profiles showed that 54% of the total Hyp had polysaccharide substituents with a medium size of 20 residues (approximately 3 kD) but as large as 52 residues (approximately 8 kD; Zhao et al., 2002). The remaining Hyp residues either had oligoarabinosides (28%) or were nonglycosylated (12%–14%). This provided further support for the Hyp contiguity hypothesis, which predicts that blocks of contiguous Hyp residues, such as those occurring in extensins (Ser-Pro1,3), are arabinosylated with oligoarabinosides, whereas noncontiguous Hyp residues, such as those occurring in AGPs (Ala-Hyp, Ser-Hyp),

Figure 1. (Continued.)

kD, and three Hyp-tri-arabinosides (1.32 kD). The glucuronorhamnoarabinogalactan has a galactan backbone with GlcpA, Rhap, and Araf side chains similar to that shown in the AG schematic in A. This model is from Qi et al. (1991). C, Primary structure of a representative Hyp-AG polysaccharide (AHP-1) released by base hydrolysis from a synthetic AGP (Ala-Hyp)51 from tobacco BY2 cells. This model is modified from Tan et al. (2004). D, Larch AG structure. R’ = dp > 4 is an undefined AG oligosaccharide. This model is modified from Ponder and Richards (1997).
are exclusively substituted with AG polysaccharide chains. Where AGP protein backbones contain both contiguous and noncontiguous Hyp residues such as gum arabic AGP (GAGP; Fig. 1B) and the Gal-rich styler glycoprotein (GaRSGP) from Nicotiana alata stylum transmitting tract tissue (Sommer-Knudsen et al., 1996), then both type II AG polysaccharide chains and short Ara oligosaccharides are found. Heterologous expression of AGPs in tobacco BY2 cells also provided the first complete sequence of an AG chain on a synthetic peptide. Alkaline hydrolysis of the purified AGP-GFP fusion protein was used to generate a population of Hyp-AGs ranging in size from 13 to 26 sugar residues, with a median size of 15 to 17 residues (Tan et al., 2004). AG-Hyp was purified and the complete primary sequence of the 15-residue glycan chain was established (Fig. 1C), confirming the generally held view of AG structure previously proposed (Fig. 1, A and B). Interestingly, the size of the AG chains elaborated on these synthetic peptides is low compared with even the smallest predicted chain length of native AGs (Fincher et al., 1983; Gane et al., 1995), and the periodate-resistant (1-3)-β-galacto-oligosaccharide “repeats” are much shorter than that estimated from studies on native AGPs (Churms et al., 1981, 1983; Bacic et al., 1987).

The type II AG chains, however, are not restricted to classical AGPs and are also found (1) on chimeric AGPs (e.g. GaRSGP; Sommer-Knudsen et al., 1996) and the chimeric, hybrid AGP 120-kD glycoprotein also from the styles of N. alata (Lind et al., 1994; Schultz et al., 1997; Lee et al., 2008); (2) covalently linked to pectins in Angelica acutiloba root walls (Yamada, 1994), in spent hops (Humulus lupulus) extracts (Oosterveld et al., 2002), and in carrot (Daucus carota) roots (Immerzeel et al., 2006); and (3) as pure polysaccharides with no detectable protein in the secretions from western larch (Larix occidentalis; Ponder and Richards, 1997; Fig. 1D).

In contrast to the lack of well-characterized amino acid sequence motifs defining O-glycosylation, N-glycosylation, when present, only occurs when the protein backbone contains the predicted conserved sequence Asn-Xaa-Ser/Thr, where Xaa can be any amino acid except Pro (Lerouge et al., 1998). Classical AGP protein backbones do not contain this consensus sequence, but many chimeric AGPs, including fasciclin-like AGPs (FLAs), GaRSGP, and an early nodulin-like protein (Johnson et al., 2003a), contain the consensus sequence, and N-glycosylation has been experimentally demonstrated for both GaRSGP (Sommer-Knudsen et al., 1996) and AtFLA7 (Johnson et al., 2003a).

Protein

Cloning of genes encoding AGP backbones has significantly advanced our understanding of AGPs. The first protein backbone genes identified by traditional biochemical approaches required the purification of native AGPs, their subsequent deglycosylation, and Edman degradation N-terminal sequencing (Chen et al., 1994; Du et al., 1994). More recently, whole genomes have been sequenced, revealing the incredible diversity of AGP protein backbones. For example, Arabidopsis has more than 100 genes encoding proteins predicted to have AGP glycomodules. The first thorough bioinformatic analysis of the Arabidopsis genome identified 47 different genes in four classes: 13 classical AGPs, three Lys-rich AGPs, 10 AG peptides, and 21 FLAs (Schultz et al., 2002). A complementary approach, looking for genes encoding GPI-anchored proteins, identified approximately 40 more genes that encoded chimeric proteins containing small AG glycomodules, including proteins such as lipid transfer proteins and selected receptor-like kinases (Borner et al., 2002).

Classical AGPs and AG peptides can be considered the basal form of AGPs in that they have no other domains that might confer functions; as such, the entire protein backbone acts as a glycosylation scaffold. Most classical AGPs have protein backbones that are approximately 100 amino acid residues, whereas most AG peptides have protein backbones of 10 to 13 amino acid residues (Schultz et al., 2002). Purification and protein sequencing of eight of 12 AG peptides from Arabidopsis confirmed prolyl hydroxylation and GPI anchoring of AG peptides and indirectly confirmed AG glycosylation, as the AG peptides could be precipitated with Yariv reagent (Schultz et al., 2004).

FLAs are a large class of chimeric AGPs in Arabidopsis and contain one or two fasciclin domains, thought to be important for protein-protein interactions, and one or two AG domains (Johnson et al., 2003a). Large gene families of FLAs exist in other dicots, such as cotton (Gossypium hirsutum; Huang et al., 2008), monocots (rice [Oryza sativa] and wheat [Triticum aestivum]; Faik et al., 2006), and gymnosperms (pine [Pinus species]; Li et al., 2009). At least two FLAs (each with only a single fasciclin domain) are found in the moss Physcomitrella patens (Lee et al., 2005). In trees such as eucalypts (Eucalyptus species) and poplars (Populus species), selected FLAs are preferentially expressed in tension wood, suggesting that they contribute to branch strength (Andersson-Gunnerås et al., 2006; Qiu et al., 2008).

The finding that AG glycomodules are found in combinations with other functional domains leads to the question: what is the function of the glycomodule in chimeric AGPs? The range of AG peptides in a few species illustrates the diversity of chimeric AGPs that is being revealed as more cDNA and gene sequences become available. In Arabidopsis, there is an AG peptide, AtAGP24, that has a putative metal-binding domain at the N terminus followed by a short AG glycomodule (HEGGHIHHAOAOOGOAS-GPI anchor, where O is Hyp; Schultz et al., 2004). A search for orthologs identified an EST in barley (Hordeum vulgare; TC147437) that has 12 HXXXH motifs interspersed with short AG glycomodules and no other domains (W.M.A. Wankamaruddin, J.M. Preuss, M.A. Ellis et al.)
GPI Anchor

A bioinformatic search of the Arabidopsis genome reveals that approximately 248 proteins, belonging to diverse classes of cell surface (glyco)proteins, including enzymes, contain the signal sequence for the addition of a GPI anchor, and approximately 40% of these contain AG glycomodules (Börner et al., 2002, 2003). Most AGP protein backbones contain the expected signals (an N-terminal secretion signal and a C-terminal hydrophobic signal sequence) specifying the cotranslational addition of a GPI anchor (Eisenhaber et al., 2003b; Johnson et al., 2003b). The C-terminal signal sequence contains small aliphatic amino acids at the ω (cleavage) and ω+2 sites, followed by a short spacer, and terminates in a stretch of small hydrophobic amino acids. The structure of the lipid moiety of the GPI anchor of two AGPs has been experimentally determined and is a phosphoceramide (Fig. 1A; Oxley and Bacic, 1999; Svetek et al., 1999). The glycan moiety of one AG GPI anchor has been fully sequenced and contains the conserved minimal structure found in all eukaryotic GPI anchors [α-1→2α-1→6α-1→4α-1→ρ-GlcNH$_2$-inositol] with a plant-specific substitution of a β-1→4Galp(1-4) residue on the third Manp (Fig. 1A; Oxley and Bacic, 1999). This remains the only plant GPI anchor sequenced to date. In animals, prolyl 4-hydroxylase (P4H; EC 1.14.11.2) is an α$\cdot$β$_2$ tetramer that is responsible for hydroxylation of Pro residues in proteins such as collagen (Hill et al., 2000), although Hyp is not glycosylated in animals. The β-subunit is a multifunctional enzyme, protein disulfide isomerase. The α-subunit is the catalytic subunit and is referred to as P4H1. In plants, early biochemical analysis suggested that only the catalytic α-subunit was required for modification of Pro to Hyp (Chrispeels, 1984; Bolwell et al., 1985). Sequencing of the Arabidopsis genome and subsequent testing of heterologously expressed candidate genes using synthetic peptides confirmed the activity of two en-

GPI anchor sequenced to date. This remains the only plant

Tertiary Structure

An understanding of the tertiary structure of AGPs will ultimately contribute to an elucidation of their molecular mechanism(s) of action in planta and explain their unique physicochemical properties of having extremely low viscosity at high concentration and a capacity to act as emulsifiers. Until recently, two models have been proposed, the “wattle blossom” model (Fincher et al., 1983), based upon their biophysical properties (Fig. 1A), and the “twisted hairy rope” model (Fig. 1B; Qi et al., 1991), for a GAGP based upon a combination of transmission electron microscopy (TEM) imaging (Fig. 2B) and chemical properties. The twisted hairy rope model envisages an alignment (and stabilization via hydrogen bonding) of the AG chains along the long axis of the protein backbone, forming a molecule of approximately 5 nm diameter, and provides an explanation of how some AGPs can permeate the pores (4–5 nm) of primary walls by reptation. Using direct TEM imaging (Fig. 2A) of an AGP from carrot suspension cultures revealed ellipsoidal rather than rod-like structures that were more consistent with the wattle blossom model (Baldwin et al., 1993). Computer-based molecular modeling of NMR data and compositional studies of a synthetic AGP from tobacco BY2 cells, of three AG chains linked to a (A-O/ P)$_6$, revealed that the bulky AG side chains do indeed hydrogen bond with the polypeptide, but in a more complex manner than originally proposed (Fig. 2, D and E). Gum arabic is generally considered to comprise three fractions: two major ones comprising an AG peptide (88.4%, generally referred to as AG and F1 [for fraction 1]), an AGP fraction (10.4%), and a third minor glycoprotein (GAGP) fraction (1.2%; Sanchez et al., 2008). The twisted hairy rope model was derived from studies of this latter GAGP fraction (Figs. 1B and 2B). In 2008, independent studies of the other (major) gum arabic fractions concluded that the AGP fraction was consistent with the wattle blossom model (Mahendran et al., 2008), whereas the AG peptide F1 was a thin oblate ellipsoid that led to the formulation of a “disc-like model” (Fig. 2C), which was difficult to reconcile with the wattle blossom model (Sanchez et al., 2008). The latter study used a combination of small angle neutron scattering coupled to ab initio calculations and various microscopic techniques (TEM and atomic force microscopy). Therefore, it appears that the three fractions from gum arabic have different tertiary structures that undoubtedly reflect the enormous primary structural diversity of both the AG chains and the protein backbones of these AGPs.

BIOSYNTHESIS

Pro Hydroxylation

In animals, prolyl 4-hydroxylase (P4H; EC 1.14.11.2) is an α$\cdot$β$_2$ tetramer that is responsible for hydroxylation of Pro residues in proteins such as collagen (Hill et al., 2000), although Hyp is not glycosylated in animals. The β-subunit is a multifunctional enzyme, protein disulfide isomerase. The α-subunit is the catalytic subunit and is referred to as P4H1. In plants, early biochemical analysis suggested that only the catalytic α-subunit was required for modification of Pro to Hyp (Chrispeels, 1984; Bolwell et al., 1985). Sequencing of the Arabidopsis genome and subsequent testing of heterologously expressed candidate genes using synthetic peptides confirmed the activity of two en-

Tester, A.A.T. Johnson, and C.J. Schultz, unpublished data). A similar contig is found in closely related wheat (TC252449), but the apparent rice ortholog (OJ990528.30.2) has the same alternating domain structure, although the sequences between AG glycomodules contain fewer His residues. A gene for a different chimeric AG peptide was recently discovered in Medicago truncatula and seems to be restricted to legumes. The predicted mature protein backbone has an N-terminal domain of unknown function of 21 amino acid residues, followed by an AG glycosylation motif APAPTP (Schultz and Harrison, 2008). This AG peptide, MtAMA1, is only expressed in root tissue colonized with symbiotic arbuscular mycorrhizal fungi, specifically in arbuscule-containing cortical cells, which are the major site of nutrient transfer in this widespread symbiosis. Differences among species highlight the evolutionary complexity of the AG peptide subfamily of AGPs and make functional studies of chimeric AGPs an even greater challenge.
coded proteins, P4H-1 (Hieta and Myllyharju, 2002) and P4H-2 (Ttainen et al., 2005). The synthetic peptides used in these early studies were more like mammalian collagen and extensin-type motifs ([Ser/Thr]-Pro$_3$) rather than the scattered ([Ala/Ser/Thr]-Pro) motifs of AGPs, so while the current knowledge of the protein code for Pro hydroxylation has provided general rules, there are still some ambiguities. For example, why is the Pro in Gly-Pro hydroxylated in the AG peptide, AtAGP24 (Schultz et al., 2004), but not in gum arabic, where Gly-Pro is frequently followed by His (Goodrum et al., 2000) rather than Ala, as in AtAGP24? The first rules suggested by Kieliszewski and Lamport (1994), namely that Lys-Pro, Tyr-Pro, and Phe-Pro are never hydroxylated, whereas Pro-Val is always hydroxylated, remain valid. The importance of the residues on either side of the Pro residue is supported by the recent solving of the crystal structure of the algal P4H from Chlamydomonas rheinhardtii (Koski et al., 2009). The active site of CrPH4-1 contains two loops that bind intimately the Ser-Pro-Ser of peptide (Ser-Pro)$_5$.

The code for hydroxylation was further extended using recombinant sporamin expressed in tobacco BY2 cells. Sporamin is a storage protein from sweet potato (Ipomoea batatas) and contains a single Pro residue that is hydroxylated and glycosylated. Mutation and deletion studies revealed a motif of seven residues ([not basic]-[not T]-[AVSG]-Pro-[AVST]-[GAVPSTC]-[APS]), suggesting a much larger active site (Shimizu et al., 2005). A rather unusual motif, KSPKKS, is hydroxylated in the 120-kD glycoprotein from N. alata (Schultz et al., 1997), highlighting the need to sequence a broad range of native and artificial AGP and extensin substrates before the code can be fully elucidated. The current approach of using plant cell suspension cultures is extremely informative but time-consuming, and downstream analysis of hydroxylation and glycosylation is difficult. It should be possible to increase the throughput of investigating Pro hydroxylation of plant AGP protein backbones using yeast-based systems that have been successful for studying mammalian P4H specificity (Vuorela et al., 1997).

The in vivo location of P4H in plants remains unclear. A recent report suggests that the enzymes are located in both the endoplasmic reticulum (ER) and the Golgi (Yuasa et al., 2005). However, the predicted Golgi localization signals are in the N-terminal ER signal sequence that would be cleaved as the proteins enter the ER cotranslationally and, therefore, would not be available for Golgi targeting.

![Figure 2](image-url)

**Figure 2.** A composite of images of AGPs purified from various plants. A, Stereopair of TEM micrographs showing the three-dimensional structure of a large aggregate of AGP from carrot suspension cultures imaged by the fast-freeze, deep-etch, rotary-shadowed replica technique. Bar = 200 nm. This image is from Baldwin et al. (1993). B, TEM micrograph of Superose-purified GAGP from gum arabic after rotary shadowing. Bar = 100 nm. This image is from Qi et al. (1991). C, TEM micrograph of the AG peptide F1 from acacia gum. The inset shows a magnification of part of the micrograph. This image is modified from Sanchez et al. (2008). D and E, Space-filling CPK models of three AHP-1 AGs (15 glycan residues per polysaccharide chain) from tobacco BY2 cells glycans linked to the C-4 of each Hyp (O) residue in the 12-residue peptide (A-P-A-O-A-O-A-O-A-P-A-P), in which each O has an AHP-1 substitution (underlined). Nitrogen atoms are shown in dark blue; the oxygen atoms are red; hydrogen atoms are gray; and carbon atoms are turquoise blue. These images are from Tan et al. (2004). D, Glycosylated (A-O/P)$_6$ (see above). Side view of a polysaccharide cluster. Three AHP-1 glycans labeled A to C are O-linked to the Hyp residues of the glycosylated (A-O/P)$_6$ model (residues 4, 6, and 8 of the peptide). The protein backbone lies across the figure, with the N terminus at the far left. Note the close proximity of polysaccharide B (green) to the polypeptide backbone, where the Ara disaccharide residues C$_1$ and C$_2$ form three H bonds as follows: the hydroxymethyl (C-5) of Ara residue C$_1$ to both the carbonyl of Hyp residue 4 and the peptide N of Ala residue 5; and the C-2 hydroxyl of Ara residue C$_2$ to the NH of Ala residue 6. In contrast, the Ara trisaccharide residues at the tip of each polysaccharide form peripheral hook-like projections; these may result in multiple weak interactions (“molecular Velcro”) with the Yariv reagent, which specifically interacts with AGPs. E, Glycopeptide (A-O/P)$_6$. End-on view of a polysaccharide cluster. Reorienting the polypeptide so that it is perpendicular to the plane of the paper shows a syndiotactic propeller-like arrangement of the AG polysaccharides around the polypeptide, providing surfaces for interactions and interdigitation with other matrix molecules.
from the ER. It is also possible that Golgi localization was an artifact of the BY2 GFP fusion protein expression system. N-terminal sequencing of P4H native proteins is needed to determine if the Golgi signal is retained in any of the mature P4H enzymes.

### The Plant GPI Anchor Pathway

Most of our knowledge of the assembly of the GPI anchor is gleaned from the exhaustive analysis performed in mammalian, yeast, and protozoan systems (for review, see Orlean and Menon, 2007; Fujita and Jigami, 2008; Kinoshita et al., 2008). Given the importance of GPI anchors in eukaryotic systems, it is surprising that there has been little progress in elucidating this pathway in plants. Nonetheless, it is highly likely that the process is conserved, as orthologs of key GPI biosynthetic genes are found in plant genomes (Fig. 3; Table I; Schultz et al., 1998). Lalanne et al. (2004) demonstrated that disruptions in the SETH1 and SETH2 genes that encode orthologs of PIG-C and PIG-A, respectively, two components of the GPI-GlcNH₂ complex, specifically block male transmission and pollen function in Arabidopsis. To

**Figure 3.** Proposed mechanism for the synthesis and addition of a GPI anchor to an AGP. The addition of a GPI anchor to an AGP likely occurs in several phases, beginning with the synthesis of the GPI moiety on the cytoplasmic surface of the ER. The protein backbone is inserted into the ER cotranslationally, and eventually the two processes converge. The GPI anchor synthesis pathway shown is a composite of that established for mammals and yeast (Orlean and Menon, 2007; Kinoshita et al., 2008) based on the structure of the pear GPI anchor (Fig. 1A). A similar pathway likely exists in plants, since orthologs of the mammalian and yeast genes have been found in plants (Table I). Proteins involved in the biosynthesis pathway (numbered) are summarized in Table I. In both mammals and yeast, any branching glycosylation of the trimannosyl core occurs between steps 8 and 10, and although galactosylation is plant specific, it is presumed to occur at the same stage (step 9). Since substitution of inositol (step 6) and the trimannosyl core (step 11) with acyl and ethanolamine phosphate residues, respectively, has not been demonstrated in plant GPIs, these steps are not shown here, although, interestingly, BLAST searches have identified putative acyl transferases (Table I). Following concomitant removal of the C-terminal GPI signal sequence and addition of the GPI anchor (step 12), numerous Pro residues present within the AGP backbone are hydroxylated to Hyp. Substitution of these Hyp residues with AG chains probably begins within the ER and is completed within the Golgi network. AGPs are transported to the plasma membrane via vesicular transport, where they are either temporarily anchored to the plasma membrane before being released by phospholipases (GPI-PLC and GPI-PLD) or endocytosed. This model is modified from Schultz et al. (1998).
our knowledge, this is the only study of GPI assembly genes in plants.

In brief, the addition of a GPI anchor to a protein occurs in several phases, beginning with the early steps in the synthesis of the GPI moiety on the cytoplasmic surface of the ER and the subsequent steps occurring in the ER lumen after the flipping of the anchor from one face to the other. Proteins to be GPI anchored are cotranslationally inserted into the ER, and eventually the two processes converge, when a transamidase complex concomitantly cleaves the C-terminal peptide and transfers the GPI anchor onto the C terminus of the mature protein by a transamidation reaction (Fig. 3). The GPI-anchored protein then undergoes prolyl hydroxylation and galactosylation (see below) before transit to the Golgi, where the AG and possibly the oligoarabinoside chains are elaborated. Within the Golgi, extensive lipid remodeling occurs that is critical to the targeting of GPI-anchored proteins to membrane microdomains or rafts enriched in sphingolipids and sterols (Orlean and Menon, 2007; Fujita and Jigami, 2008; Kinoshita et al., 2008). It is thought that association with membrane rafts is important for GPI-anchored protein function. Lipid rafts appear to exist in plants (Borner et al., 2005), but their composition and function are mostly undefined for plants, although lipid rafts have been implicated in stabilizing cellulose synthase complexes in the plasma membrane (V. Bulone, personal communication).

### Biosynthesis of the AG and Arabino-Oligosaccharide Chains

The biosynthesis of glycoconjugates has been extensively studied in mammals, yeast, and bacteria, where many glycosyltransferases (GTs) have been cloned and biochemically characterized (for review, see Breton et al., 2002; Coutinho et al., 2003). However, only a few GTs have been purified and then cloned from plants, and the majority are involved in the assembly of cell wall polysaccharides (for review, see Doblin et al., 2010). Based upon the assembly of mammalian proteoglycans and recent molecular and biochemical studies in Arabidopsis by Strasser et al. (2007) and Qu et al. (2008), it is proposed that the AGs on AGPs are synthesized by type II ER/Golgi-located GTs, including members of the CAZy GT-family-31 GTs that possess (1,3)-β-GalT activity (for review, see Egelund et al., 2008).

### Table I. GPI biosynthesis proteins and their putative plant orthologs (after Orlean and Menon, 2007)

Steps refer to the individual processes involved in GPI anchor biosynthesis and are identical to those used in Figure 3. Putative plant orthologs were identified by tBLASTn or BLASTp using the human and yeast protein sequences to search the Arabidopsis and rice genomes. Searches were performed at the National Center for Biotechnology Information or www.gramene.org using default settings (except to find PIG-Y, where BLASTp word size was changed from n = 3 to n = 2, with no filtering). For Arabidopsis sequences, putative orthologs were cross-referenced, where available, against genes previously identified by Eisenhaber et al. (2003a). n/a, Not applicable.

<table>
<thead>
<tr>
<th>Step</th>
<th>Enzyme</th>
<th>Mammalian Protein</th>
<th>Yeast Protein</th>
<th>Putative Plant Orthologs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GPI-GlcNAc transferase</td>
<td>PIG-A</td>
<td>Gpi3p</td>
<td>SETH2, Atg45100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-C</td>
<td>Gpi2p</td>
<td>SETH1, Atg34980</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-H</td>
<td>Gpi15p</td>
<td>Atg43530</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-P</td>
<td>Gpi19p</td>
<td>Atg61280</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-Q</td>
<td>Gpi1p</td>
<td>Atg51710</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-Y</td>
<td>Eri1p</td>
<td>Atg52312 (83696...83338)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-M</td>
<td>Gpi12p</td>
<td>Atg27340</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-L</td>
<td>Gpi12p</td>
<td>Atg27340</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-W</td>
<td>Gwt1p</td>
<td>Atg17910</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-X</td>
<td>Pbn1p</td>
<td>Atg46850</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-N</td>
<td>Mcd4p</td>
<td>Atg61830</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-V</td>
<td>Gpi18p</td>
<td>Atg11880</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-B</td>
<td>Gpi10p</td>
<td>Atg14850</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-O</td>
<td>Gpi13p</td>
<td>Atg17250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-F</td>
<td>Gpi11p</td>
<td>Atg16040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-G</td>
<td>Gpi7p</td>
<td>Atg22530</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-U</td>
<td>Gpi11p</td>
<td>Atg16040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-K</td>
<td>Gpi8</td>
<td>Atg8750</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GAA1</td>
<td>Gaa1p</td>
<td>Atg19130</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GAA1</td>
<td>Gaa1p</td>
<td>Atg682200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GAA1</td>
<td>Gaa1p</td>
<td>Atg682200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-T</td>
<td>Gpi16p</td>
<td>Atg30710</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PIG-U</td>
<td>Gah1p</td>
<td>Atg63110</td>
</tr>
</tbody>
</table>

*The Arabidopsis sequence is not annotated but does contain significant sequence similarity to the rice Os01g0939000 (63% identity, 80% similarity) putative ortholog if the ER signal sequence is removed.*

Ellis et al.
et al., 2010). The O-galactosyltransferase, which is the initial enzyme in the AG biosynthetic pathway (adding the first Gal residue to a Hyp residue in the protein backbone), is predominantly located in ER fractions after Suc density-gradient centrifugation of extracts from Arabidopsis T87 cells (Oka et al., 2010). A model has been proposed in which the addition of the first Gal onto Hyp occurs in the ER (Oka et al., 2010). Early studies by Fincher and colleagues (Mascara and Fincher, 1982; Schibeci et al., 1984) showed that the enzymes responsible for the synthesis of polymers containing (1,6)\-Gal linkages in Italian ryegrass (Lolium multiflorum) were associated exclusively with the subcellular fractions enriched in Golgi-derived membranes and therefore concluded that the Golgi apparatus plays an important part in the synthesis of the carbohydrate component of AGPs. Although 30 years have passed since these early studies, none of the enzymes responsible for the synthesis of the Gal chains have been cloned and characterized, nor have their substrate specificities been determined biochemically. It is uncertain, therefore, whether the addition of the remaining sugars that make up the AG chains occurs one residue at a time or blockwise, as is the case for N-linked glycans (Kelly et al., 2006), although recent results (Qu et al., 2008; Oka et al., 2010) support the former mechanism. Therefore, it is reasonable to assume that for the assembly of AG chains, several GalTs will be required, such as O-galactosyltransferases, (1,3)\-GalTs, and (1,6)\-GalTs, and that these enzymes will work coordinately to regulate the density, length, and sequence of AG chains.

Several GTs are predicted to be responsible for decorating the termini of AG chains (AraTs/GlcATs/RhATs). Fuc residues are present in AGPs in several dicot plants as terminal residues linked via an \(\alpha\)-linkage to the C-2 position of an adjacent Araf residue (Tsumauraya et al., 1988). Two members of CAZy GT-family-37, AtFUT4 and AtFUT6, were recently characterized as \(\alpha\)(1,2)-fucosyltransferases (FUTs) that are specific for AGPs (Wu et al., 2010). These are the first enzymes to be characterized that are specific for AGP glycosylation. The authors demonstrated that these two enzymes differentially fucosylate AGPs at different positions on the AG side chains, indicating that the two FUT enzymes may have different physiological roles. An AtFUT6-GFP chimera was localized to the Golgi apparatus, supporting the evidence that many of the biosynthesis of the AG side chains occurs within this organelle (Wu et al., 2010).

The assembly of the arabino-oligosaccharides remains largely unstudied. Two putative Arabidopsis AraT insertion mutants, \(rra1\) and \(rra2\), are presumed to be involved in extensin arabino-oligosaccharide assembly (Egelund et al., 2007) based on reduced residual Ara content in a polymer tightly associated with the cellulosic wall residue. These AraTs possess a type II GT topology and are likely Golgi located (J. Egelund, unpublished data), but whether this reflects the location and activity of the AraTs involved in the synthesis of AGPs remains to be demonstrated. An interesting new approach from the Pauly laboratory (Gille et al., 2009) used glycosyl hydrolases as a forward chemical genetic screen to identify cell wall mutant lines where extensin arabinosylation was altered. It will be of interest to see if any of the other 22 genetic loci identified in this screen uncover AGP-specific GTs or whether it will be necessary to use AGP-specific hydrolases in a future screen.

FUNCTIONS

Biological

The function of AGPs was recently summarized in an excellent review by Seifert and Roberts (2007). Thus, this section will outline key points about AGP function with regard to the GPI anchor, AG side chains, and the fully glycosylated mature protein, highlighting recent studies and identifying future strategies to further clarify AGP function(s).

The function of AGPs has been hard to ascertain for many reasons. The heterogenous nature of the AGP family suggests that AGPs do not have one specific role, as would be anticipated from studies of mammalian glycoproteins/proteoglycans (Filmus et al., 2008; Schaefer and Schaefer, 2010). The complex nature of the carbohydrate side chains buries the protein backbone, preventing simple isolation of individual AGPs, and in many cases prevents detection by protein-specific antibodies. The complex glycosylation and high degree of hydroxylation of AGP protein backbones does not allow expression of correctly glycosylated recombinant proteins in microorganisms, which would allow in vitro examination of function. The large size of the AGP gene family likely leads to functional redundancy, and the production of double mutants of homologous AGPs is often required to produce phenotypes (Motose et al., 2004; Coimbra et al., 2009). For other subclasses of AGPs, such as the three Lys-rich AGPs of Arabidopsis, AtAGP17, AtAGP18, and AtAGP19, each family member may have a unique function based on different phenotypes. It is not known if these differences are due solely to their different expression patterns or whether it is also due to different interacting partners or downstream signaling events (Acosta-Garcia and Vielle-Calzada, 2004; Gaspar et al., 2004; Yang et al., 2007).

Despite these limitations AGPs have been implicated in many processes involved in plant growth and development, including somatic embryogenesis (van Hengel et al., 2002), root growth and development (van Hengel and Roberts, 2003), hormone responses (Park et al., 2003), signaling (Schultz et al., 1998), xylem differentiation (Motose et al., 2004), resistance to Agrobacterium tumefaciens-mediated infection (Gaspar et al., 2004), initiation of female gametogenesis (Acosta-Garcia and Vielle-Calzada, 2004), salt tolerance (Shi et al., 2003; Lamport et al., 2006), cell wall plasticizer (Lamport et al., 2006), and seed coat development (van Hengel et al., 2002), to name a few (Table 1). In this section we will concentrate on the role of AGPs in cell wall plasticization and on strategies to further clarify AGP function(s).
cell expansion (Lee et al., 2005; Yang et al., 2007), secretion (Xu et al., 2008a), promotion of pollen tube growth and guidance (Cheung et al., 1995; Wu et al., 1995, 2000; Mollet et al., 2002; Lee et al., 2008), programmed cell death (Gao and Showalter 1999; Chaves et al., 2002), pollen grain development (Pereira et al., 2006; Levitin et al., 2008; Coimbra et al., 2009), and self-incompatibility in pollen (Lind et al., 1996; Cruz-Garcia et al., 2005; Hancock et al., 2005; Lee et al., 2008; Table II).

Studies of AGPs often use methods that provide a global approach for determining AGP function. Biochemical analysis of AGPs utilizes carbohydrate-specific monoclonal antibodies for which precise epitope specificities are often unknown (for review, see Knox, 1997). \( \beta \)-Gal Yariv, the pink/orange dye that specifically binds AGPs, is used in many studies to bind AGPs, interfering with the activity of all AGPs and thus indirectly elucidating function. Both \( \beta \)-Yariv and the carbohydrate-directed monoclonal antibodies are estimated to bind 50 to 100 different AGP proteins and thus provide a global view of AGP distribution, localization, and function.

The use of molecular techniques has greatly expanded our ability to study individual AGPs directly. These include transcript analysis and the production of mutant plant lines, such as T-DNA insertion lines, RNA interference (RNAi) lines, and lines overexpressing individual genes. These studies are summarized in Table II. This approach is not without difficulties, as, in some cases, a phenotype is only apparent after the production of double knockout lines (Motose et al., 2004), and in others, conditional stress, such as salt, is required to induce a phenotype (Shi et al., 2003). Studies of the same gene by different methods highlight the complex nature of AGP expression. Tissue- and growth stage-specific transcript analysis often does not concur with localization using AGP promoter-driven constructs (Acosta-Garcia and Vielle-Calzada, 2004; Yang and Showalter, 2007), suggesting that multiple approaches are required. For example, a T-DNA insertion in the promoter region of the \( \text{rat1} \) mutant line specifically inhibits AtAGP17 expression in roots but not leaves (Gaspar et al., 2004), indicating the importance of the promoter and untranslated regions in controlling AGP expression.

The analysis of RNA transcripts allows us some insights into AGP distribution within plants. In many cases, a clear expression pattern both in tissue location and developmental stage is evident for many individual AGPs (summarized in Table II). For many individual AGPs, transcription occurs at a higher level in specific plant tissues such as roots and root tips (van Hengel and Roberts, 2003), flowers (Acosta-Garcia and Vielle-Calzada, 2004), pollen tubes and pollen (Pereira et al., 2006), young stems (Gilson et al., 2001), differentiating tissue (Motose et al., 2004; Dahiya et al., 2006), shoot apices (Park et al., 2003), and filament apices (Lee et al., 2005). In other cases, specific AGPs are found in several tissue types within the plant; for example, \( \text{SOS5/FLA4} \) is expressed ubiquitously throughout the plant but mutant plants only show aberrant root morphology during salt stress (Shi et al., 2003). AGPs have a key function in young or rapidly expanding or elongating tissues and during times of stress.

The presence of a GPI anchor is important when considering AGP function. GPI anchors do not traverse the entire plasma membrane, and in mammalian systems GPI-anchored proteins have been implicated in mediating cell-cell interactions by interacting with other proteins that possess transmembrane domains (Schultz et al., 1998; Johnson et al., 2003b; Filmus et al., 2008; Lim et al., 2008). It is unlikely that direct cell-to-cell interactions occur between GPI-anchored proteins in adjacent plant cells as occurs in mammalian cells, due to the thickness of the wall, estimated to be around 100 nm for a primary wall (Lamport et al., 2006). However, GPI-anchored AGPs and/or their soluble forms could interact with plasma membrane-bound receptor kinases or soluble forms of AGPs could interact with receptors in neighboring cells. The FLAs are expected to be involved in such interactions, as they contain fasciclin domains (Johnson et al., 2003a), which in other eukaryotes are involved in cell adhesion as a result of protein-protein interactions (Elkins et al., 1990; Kawamoto et al., 1998). Recently, it was proposed that \( \text{SOS5/FLA4} \) is a ligand of two members of the Leu-rich repeat receptor-like kinase family, FEI1 and FEI2, based on genetic analyses indicating that FEI1 and FEI2 act within the same pathway as \( \text{SOS5/FLA4} \) (Xu et al., 2008b).

Localization of GPI-anchored LeAGP1-GFP to both the plasma membrane and Hechtian strands in tobacco suspension cultures demonstrates a link between AGPs, the plasma membrane, and the cytoskeleton (Sun et al., 2004; Sardar and Showalter, 2007). Classical GPI-anchored AGPs have been implicated in controlling cell shape via a connection with the cytoskeleton, due to their involvement in orienting cortical microtubules and influencing F-actin polymerization (Andémé-Onzighi et al., 2002; Sardar et al., 2006; Nguema-Ona et al., 2007). It has also been proposed that GPI-anchored AGPs may act as a plasticizer to loosen the pectin network or as a stabilizer of the wall during times of cell stress or expansion (Lamport et al., 2006).

Several studies suggest that the glycan chains are important for AGP function. Addition of a single AG glycosylation site to human growth hormone increased its secretion 100-fold from tobacco suspension cells (Xu et al., 2008a). Evidence for the importance of the carbohydrate chains in determining biological function comes from the \( \text{mur1} \) and \( \text{reb1} \) mutants. The \( \text{mur1} \) mutants are defective in the biosynthesis of Fuc and have aberrant AGP structures and altered root morphology, which is not attributed to defects in other wall components (van Hengel and Roberts, 2002). The \( \text{reb1} \) mutant plants, defective in Gal biosynthesis (Seifert et al., 2002), have abnormal root extension,
Table II. Summary of recent functional studies of AGPs

Only studies using forward or reverse genetics providing plant lines with definable phenotypes are listed.

<table>
<thead>
<tr>
<th>Name</th>
<th>AGP Class</th>
<th>GPI</th>
<th>Location?</th>
<th>Genetic Analysis</th>
<th>Phenotype/Observed Result</th>
<th>Proposed Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PpAGP1</td>
<td>Classical</td>
<td>Yes</td>
<td>Filament apices</td>
<td><strong>agp1 knockout line</strong></td>
<td>Reduced colony size due to decreased cell length</td>
<td>Apical cell expansion</td>
<td>Lee et al. (2005)</td>
</tr>
<tr>
<td>AtAGP6</td>
<td>Classical</td>
<td>Yes</td>
<td>Pollen and pollen tubes Pollen grain and pollen tube</td>
<td><strong>agp6 agp11 double T-DNA insertion mutants, double RNAi knockdowns, AGP6 point mutant</strong></td>
<td>Double insertion mutants show collapsed pollen grains; RNAi double knockdowns have reduced pollen tube length, shorter siliques, and fewer seeds</td>
<td>Pollen grain development and pollen tube growth</td>
<td>Levitin et al. (2008); Coimbra et al. (2009)</td>
</tr>
<tr>
<td>AtAGP11</td>
<td>Classical</td>
<td>Yes</td>
<td>Pollen grain and pollen tube</td>
<td><strong>rat1 T-DNA insertion in promoter (specific inhibition of transcription in the roots)</strong></td>
<td>Reduced Agrobacterium binding to the root surface</td>
<td>Resistance to infection</td>
<td>Nam et al. (1999); Gaspar et al. (2004)</td>
</tr>
<tr>
<td>CsAGP1</td>
<td>Lys-rich</td>
<td>Yes</td>
<td>Shoot apices, roots, hypocotyls, and cotyledons Maturing metaxylem elements of young stems and petioles, transmitting tissue of cells of the style, inner and outer phloem, cambial zone of young stems and petioles</td>
<td>Overexpression in tobacco plants</td>
<td>Tobacco plants overexpressing CsAGP1 are taller with longer internodes</td>
<td>Stem elongation, hormone response</td>
<td>Park et al. (2003)</td>
</tr>
<tr>
<td>LeAGP1</td>
<td>Lys-rich</td>
<td>Yes</td>
<td>Maturing metaxylem elements of young stems and petioles, transmitting tissue of cells of the style, inner and outer phloem, cambial zone of young stems and petioles</td>
<td>Overexpression of GFP-LeAGP1 fusion protein in tomato plants</td>
<td>Reduced stem elongation, increased lateral branches, reduced seed size and less seed produced; transgenic plants that produce a non-GPI-anchored form of GFP-LeAGP1 or are missing the lys-rich domain have wild-type phenotype</td>
<td>Plant growth and development</td>
<td>Gao et al. (1999); Sun et al. (2004)</td>
</tr>
<tr>
<td>AtAGP17</td>
<td>Lys-rich</td>
<td>Yes</td>
<td>Flower, leaves, and roots</td>
<td><strong>rat1 T-DNA insertion in promoter (specific inhibition of transcription in the roots)</strong></td>
<td>Smaller, rounder rosette leaves, lighter green leaves containing less chlorophyll; delayed growth; shorter hypocotyls and inflorescence stems, fewer siliques, and less seed</td>
<td>Plant growth and development, including cell division and expansion, leaf development, and reproduction</td>
<td>Yang et al. (2007)</td>
</tr>
<tr>
<td>AtAGP19</td>
<td>Lys-rich</td>
<td>Yes</td>
<td>mRNA abundant in stems, moderate levels in flowers and roots, lower levels in leaves</td>
<td>Null T-DNA insertion line</td>
<td>Smaller, rounder rosette leaves, lighter green leaves containing less chlorophyll; delayed growth; shorter hypocotyls and inflorescence stems, fewer siliques, and less seed</td>
<td>Plant growth and development, including cell division and expansion, leaf development, and reproduction</td>
<td>Yang et al. (2007)</td>
</tr>
</tbody>
</table>

(Table continues on following page.)
<table>
<thead>
<tr>
<th>Name</th>
<th>AGP Class</th>
<th>GPI</th>
<th>Location*</th>
<th>Genetic Analysis</th>
<th>Phenotype/Observed Result</th>
<th>Proposed Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtAGP18</td>
<td>Lys-rich</td>
<td>Yes</td>
<td>Reproductive tissue</td>
<td>RNAi lines</td>
<td>Reduced seed set, and the defect is female specific; no female gametophyte is produced because the megaspore fails to enlarge and divide by mitosis</td>
<td>Initiation of female gametogenesis</td>
<td>Acosta-Garcia and Vielle-Calzada (2004)</td>
</tr>
<tr>
<td>AtAGP24</td>
<td>AG peptide</td>
<td>Yes</td>
<td>Abscission zones of seed pods</td>
<td>Overexpression of peptide IDA (for inflorescence deficient in abscission)</td>
<td>Visible deposition of Yariv-positive AGPs at abscission zones; AtAGP24 is the only AGP gene that is overexpressed based on microarray analysis</td>
<td>Metal binding, nonenzymic cleavage of cell walls</td>
<td>Stenvik et al. (2006)</td>
</tr>
<tr>
<td>GhAGP4</td>
<td>FLA (one fasciclin domain, two AGP domains)</td>
<td>Yes</td>
<td>Cotton fibers</td>
<td>RNAi line</td>
<td>Suppression of GhAGP4 and partial suppression of three other related FLAs; inhibition of fiber initiation and elongation, significantly shorter fiber length</td>
<td>Initiation and elongation of cotton fibers</td>
<td>Li et al. (2010)</td>
</tr>
<tr>
<td>SOS5/AtFLA4</td>
<td>FLA (two fasciclin domains, two AGP domains)</td>
<td>Yes</td>
<td>Flowers, leaves, roots, stems, and siliques</td>
<td>Single-base-pair substitution in conserved fasciclin domain</td>
<td>Swollen root tips and reduced root length under increasing salt concentrations</td>
<td>Salt tolerance and normal root expansion; possible ligand or involved in the production or presentation of a ligand for FEI receptor kinases</td>
<td>Shi et al. (2003); Xu et al. (2008b)</td>
</tr>
<tr>
<td>AtFLA11, AtFLA12</td>
<td>FLA (one fasciclin domain, two AGP domains)</td>
<td>Yes</td>
<td>Inflorescence stems</td>
<td>Atfla11/12 double T-DNA insertion mutants</td>
<td>Decreased tensile strength and stiffness in stems; reduced cellulose, Ara, and Gal contents and increased lignin content in stem cell walls</td>
<td>Contributors to plant stem strength</td>
<td>MacMillan et al. (2010)</td>
</tr>
<tr>
<td>ZeXyp</td>
<td>Chimeric AGP&lt;sup&gt;b&lt;/sup&gt; (contains nonspecific lipid transfer protein domain)</td>
<td>Yes</td>
<td>Immature xylem cells in stems</td>
<td>Overexpression in tobacco suspension cultures</td>
<td>Xylem differentiation (induces tracheary element differentiation)</td>
<td>Soluble signal mediating cell-to-cell interactions in vascular development</td>
<td>Motose et al. (2004)</td>
</tr>
</tbody>
</table>

*Table continues on following page.*
bulging of root epidermal hairs, and lack AGPs in their trichoblast cells (Andéme-Onzighi et al., 2002). However, reb1-1 mutants also have defects in the xyloglucan content of their walls (Nguema-Ona et al., 2006), highlighting the difficulty of attributing the phenotype from a mutation in a major sugar pathway to a single class of molecules. The production of mutants where the AG glycans are specifically modified will add more insights into precisely how the structure of the carbohydrate effects function (Egelund et al., 2010).

Several recent mutational studies have examined the role of AGP proteins in pollen and pollen tubes. Sexual reproduction in flowering plants involves pollen grains landing on the stigma surface, germination, and growth of the pollen tube via a process of rapid tip elongation through the pistil extracellular matrix. The extracellular matrix is rich in AGP-like molecules, such as the 120-kD and Transmitting Tissue-Specific (TTS) proteins, some of which are internalized by the pollen tube (Lind et al., 1996; Lee et al., 2009).

### Table II. (Continued from previous page.)

<table>
<thead>
<tr>
<th>Name</th>
<th>AGP Class</th>
<th>GPI</th>
<th>Locationa</th>
<th>Genetic Analysis</th>
<th>PhenoType/Observed Result</th>
<th>Proposed Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtXYP1, AtXYP2</td>
<td>Chimeric AGP (contains nonspecific lipid transfer protein domain)</td>
<td>Yes</td>
<td>Immature xylem cells in stems</td>
<td>Double T-DNA insertion mutants</td>
<td>Discontinuous and thicker veins, incorrect interconnections of tracheary element</td>
<td>Soluble signal mediating cell-to-cell interactions in vascular development</td>
<td>Motose et al. (2004)</td>
</tr>
<tr>
<td>AtAGP30</td>
<td>Chimeric AGP (contains Cys-rich domain)</td>
<td>No</td>
<td>Root, root tip, embryos immediately prior to germination</td>
<td>T-DNA insertion line</td>
<td>Mutants germinate more rapidly in the presence of the germination inhibitor abscisic acid; suspension cultures of cells derived from agp30 proliferate successfully but never initiate root formation</td>
<td>Root development and growth; modulation of abscisic acid perception during seed germination</td>
<td>van Hengel and Roberts (2003)</td>
</tr>
<tr>
<td>NtTTS/NaTTS</td>
<td>Chimeric AGP (contains Cys-rich C-terminal domainb)</td>
<td>No</td>
<td>Produced in style, but mature protein taken up by pollen tubes</td>
<td>Antisense line</td>
<td>Reduced fertility and reduced rate of pollen tube growth when style TTS is reduced; antisense pollen grows normally and leads to normal seed set in antisense pollen</td>
<td>Promoting pollen tube growth and guidance</td>
<td>Cheung et al. (1995)</td>
</tr>
<tr>
<td>Na120 kD</td>
<td>Chimeric and hybridc AGP (contains contiguous and noncontiguous Pro(Hyp)-rich regions and a Cys-rich C-terminal domain)</td>
<td>No</td>
<td>Pistil, and cell wall and mature protein taken up by cytoplasm/vacuoles of pollen tubes</td>
<td>RNAi in self-compatible × self-incompatible hybrid cross</td>
<td>Plants with reduced 120-kD expression unable to perform pollen rejection</td>
<td>Role in S (self-incompatibility)-specific pollen rejection</td>
<td>Hancock et al. (2005); Lee et al. (2008)</td>
</tr>
</tbody>
</table>

aLocation refers to the expression of mRNA and/or protein as determined by a variety of methods. For details, see references cited within papers describing the genetic analysis. bChimeric AGPs have one or more non-Pro(Hyp)-rich domains in addition to an AGP domain with noncontiguous Pro(Hyp) residues. Several chimeric AGPs have a conserved C-terminal domain of approximately 140 amino acid residues that contains six conserved Cys residues. cHybrid AGPs contain, for example, both an AGP domain (noncontiguous Hyp) and an extensin-like domain (contiguous Hyp) and may also be chimeric [containing a non-Pro(Hyp)-rich domain].
In *N. alata*, fully glycosylated TTS is implicated in promoting pollen tube growth and affecting pollen tube guidance, whereas the deglycosylated form of TTS is not (Cheung et al., 1995; Wu et al., 1995, 2000). TTS and 120-kD proteins also bind to three interacting partners, a putative Cys protease, a pollen-specific C2 domain-containing protein, and a S-RNase-binding protein (Cruz-Garcia et al., 2005; Hancock et al., 2005). It is proposed that binding of these proteins to pistil AGPs may contribute to signaling and trafficking inside pollen tubes (Lee et al., 2008).

In Arabidopsis, *AGP6* and *AGP11* were shown to be specific to both pollen tubes and pollen grains (Pereira et al., 2006). Double knockout mutants show aberrant pollen development, including pollen grain collapse and retraction of the plasma membrane from pollen walls (Coimbra et al., 2009). Arabidopsis lines expressing an *agp6 agp11* double knockout (RNAi) construct reduce pollen tube growth, and fewer and shorter siliques are produced (Levitin et al., 2008). This implies that AGPs are required for the integrity of the pollen grain and pollen tube growth.

The importance of a subclass of AGPs, the FLAs, is highlighted by recent studies of cotton fiber development (Liu et al., 2008). RNAi lines suppressing *GhAGP4* (a FLA) and three other FLAs resulted in inhibition of cotton fiber initiation and reduced fiber elongation (Li et al., 2010).

The essential function of AGPs in varied roles from cell expansion, growth, and differentiation to signaling, secretion, and linking the plasma membrane with the cytoskeleton and wall is clear. However, despite numerous studies, the precise mode of action of AGPs is unknown. The heterogeneous nature of AGPs and the diversity of functions suggest that a more targeted approach is required. It seems likely that the protein backbone, GPI anchor, and AG component all play roles in AGP function, and each of these areas needs to be addressed. With regard to the protein component, a targeted approach on each AGP subfamily is needed, taking into account cell type-specific expression patterns of individual family members. For example, targeted knockout of whole classes or subclasses of AGPs (e.g. by RNAi or using artificial microRNAs; Ossowski et al., 2008) may aid in defining a specific function to a specific group but may only be applicable to recently duplicated family members such as *AtAGP6* and *AtAGP11*. Domain-swapping experiments and the identification of interacting partners are needed to identify regions of the protein backbone that are of biological significance and to indicate pathways in which the AGPs are intrinsically involved.

**CONCLUSION**

This is clearly an exciting period for researchers to enter the field of AGPs. Modern molecular and genetic techniques coupled with cell biology and biochemical methods offer the opportunity to rapidly fill in several major gaps in our knowledge of this complex family of molecules, namely unequivocal biological function, the assembly of the glycan chains, and their post-assembly processing, both at the cell surface (plasma membrane release/cell wall cross-linking) and following endocytosis (turnover).

**ACKNOWLEDGMENT**

We are grateful to Ms. Joanne Noble for administrative and editing assistance and figure designs.

Received March 8, 2010; accepted April 1, 2010; published April 13, 2010.

**LITERATURE CITED**


Arabinogalactan-Proteins: Key Regulators at the Cell Surface?


Kelly J, Jarrell H, Millar L, Tessler L, Foti LA, Paul AC, Allan B,
Arabinogalactan-Proteins: Key Regulators at the Cell Surface?


Yariv J, Lis H, Hatchalski E (1967) Precipitation of Arabic acid and some seed polysaccharides by glycosylphospholipid. Biochem J 105: 1c–2c


CORRECTIONS

Vol. 153: 1555–1562, 2010

Takemiya A. and Shimazaki K.-i. Phosphatidic Acid Inhibits Blue Light-Induced Stomatal Opening via Inhibition of Protein Phosphatase 1.

The “1” at the end of the title above was not included when this article was first published. The title in the online version of the article has been corrected.

www.plantphysiol.org/cgi/doi/10.1104/pp.110.900334

Vol. 153: 403–419, 2010

Ellis M., Egelund J., Schultz C.J., and Bacic A. Arabinogalactan-Proteins: Key Regulators at the Cell Surface?

The authors have revised the final paragraph on Page 408 of this article to better describe the work done by Yuasa et al.:

P4H in plants is a type II integral membrane protein located predominantly in the Golgi lumen, making it quite different from animal P4H enzymes, which are found in the lumen of the endoplasmic reticulum (ER). This is based on the detailed analysis of His-tagged and GFP fusion forms of P4H1 (GenBank accession no. AB119250) expressed in tobacco BY2 cells by Matsuoka and colleagues (Yuasa et al., 2005). His-tagged version of NtP4H1 was purified from BY2 cells and shown to hydroxylate a collagen-like peptide, (Pro-Pro-Gly)\textsubscript{10}, in vitro. The low efficiency of hydroxylation observed is likely due to the non-native substrate used. Immunolocalization of both the His-tagged and GFP fusion constructs confirms a Golgi location with some fusion proteins also found in the ER. Furthermore, using immunoelectron microscopy and a novel antibody preparation, Goto and colleagues have demonstrated clear Golgi location of P4H (Y. Goto, K. Toyooka, and K. Matsuoka, personal communication). These data indicate that hydroxylation of Pro residues, and subsequent O-glycosylation of these Hyp residues, occurs primarily in the Golgi.

www.plantphysiol.org/cgi/doi/10.1104/pp.110.900335